Comparing and Evaluating Interest Points

Cordelia Schmid, Roger Mohr and Christian Bauckhage
INRIA Rhône-Alpes, 655 av. de l'Europe, 38330 Montbonnot, France
Cordelia.Schmid@imag.fr

Abstract

Many computer vision tasks rely on feature extraction. Interest points are such features. This paper shows that interest points are geometrically stable under different transformations and have high information content (distinctiveness). These two properties make interest points very successful in the context of image matching. To measure these two properties quantitatively, we introduce two evaluation criteria: repeatability rate and information content.

The quality of the interest points depends on the detector used. In this paper several detectors are compared according to the criteria specified above. We determine which detector gives the best results and show that it satisfies the criteria well.

1 Introduction

Interest points are locations in the image where the signal changes two-dimensionally. Examples include corners[1] and T-junctions, as well as locations where the texture varies significantly. Figure 1 shows an example of interest points detected on the "sower" painting of Van Gogh.

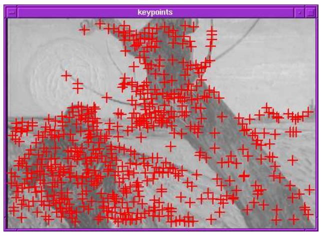

Figure 1: Interest points detected on the "sower" painting of Van Gogh. The detector used is OurHarris.

Many different interest point detectors exist in the literature. The results of different detectors vary and it is necessary to assess their performance. However, there exists no comparison of different detectors. Quantitative evaluation on real images has only been performed to assess the localization accuracy of one detector which fits a template to the signal: Brand and Mohr [1] measure the alignment of the extracted points and the precision of 3D reconstruction for their detector.

In this paper the two criteria for quantitative evaluation are repeatability rate and information content. Repeatability signifies that we obtain the points with the same pre-image independently of changes in the imaging conditions: a 3D point should be detected in both images if it is detected in one of them. The repeatability rate is the percentage of the total observed points which are repeated between two images. This is illustrated in figure 2, which shows interest points detected on the scene of figure 1 under image rotation and scale change. The repeatability rates are 92% and 71% respectively.

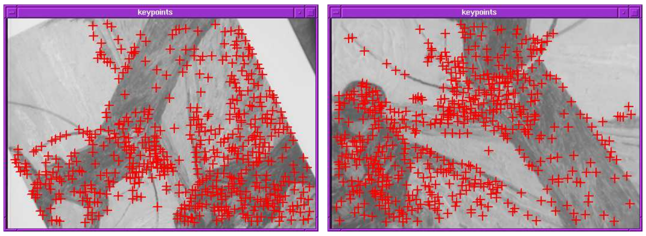

Figure 2: Interest points detected by OurHarris on the scene of figure 1. The image rotation between the left image and the one in figure 1 is 155° ; the repeatability rate is 92%. The scale change between the right image and the one in figure 1 is 1.4 ; the repeatability rate is 71%.

The second criterion, information content, measures the distinctiveness of the local greylevel pattern at an interest point. A local pattern is described using rotationally invariant combinations of derivatives. The entropy of these invariants is measured for a set of interest points.

The significance of an evaluation criterion depends on the context. Repeatability and information content are the key criteria for image matching. In this context, at least a subset of the points have to be detected at the same relative position to allow feature correspondence. In addition, interest points should have distinctive patterns to be distinguishable. Our quantitatively good results with respect to the two criteria explain why interest points have proven very successful in image matching.

The paper is organized as follows. Section 2

[1] The terms "interest points" and "corners" are often used equivalently in the literature. In this work "corner" is a special subclass of "interest point".

presents a state of the art on interest points as well as implementation details for the detectors used in our comparison. Section 3 defines the repeatability criterion, explains how to determine it experimentally and presents the results of a comparison under various image transformations. Section 4 describes the information content criterion and evaluates results for different detectors. In section 5 we select the detector which gives the best results and show that the quality of its results is very high.

2 Interest point detectors

2.1 State of the art

Interest point detectors can be divided into contour based methods, signal based methods and methods based on template fitting. Each category is presented briefly and details are given for the detectors used in our comparison. For a more complete overview the reader is referred to [11].

The idea of contour based methods is either to search for maximal curvature or inflexion points along the contour chains, or to do some polygonal approximation and to search for intersection points. Horaud et al. [7] extract line segments from the image contours. These segments are grouped and intersections of grouped line segments are the interest points.

Signal based methods compute a measure which indicates the presence of an interest point directly from the signal. It is for example possible to use the auto-correlation function of the signal. Harris [5] calculates a matrix which is related to this auto-correlation function. The squared first derivatives of the signal are averaged over a window. The eigenvalues of the resulting matrix are the principal curvatures of the auto-correlation function. If these two curvatures are high, an interest point is declared.

Cottier [2] applies the Harris detector only to contours in the image and successively uses two different window sizes to increase the location accuracy.

Förstner [4] classifies image pixels into categories - region, contour and interest point - by using the auto-correlation function. Local statistics allow a blind estimate of signal-dependent noise variance and thus an automatic selection of thresholds.

Heitger et al. [6] developed an approach inspired by experiments on the biological visual system. They extract 1D directional characteristics by convolving the image with orientation-selective Gabor filters. In order to obtain 2D characteristics, they compute the first and second derivatives of the 1D characteristics.

Methods which fit a template to the signal potentially provide sub-pixel accuracy. The image signal is fitted to a parametric model of a specific type of interest point, for example a corner or a vertex. These methods are not applicable in a general context ; we have therefore not included them into our comparison.

2.2 Implementation details

The detectors included in the comparison are Harris [5], OurHarris, Heitger [6], Förstner [4], Horaud [7] and Cottier[2]. Except in the case of OurHarris, we have used the code implemented by the original authors with a fixed standard set of parameters.

OurHarris is an improved version of Harris in which derivatives are computed more precisely. Harris convolves with the mask [-2 -1 0 1 2] to compute derivatives whereas OurHarris convolve with derivatives of the Gaussian function. A recursive implementation of the Gaussian filters [3] guarantees fast detection.

3 Repeatability

3.1 Measuring repeatability

Repeatability is defined by the image geometry. Given a 3D point P and two projection matrices M_1 and M_i, the projections of P into images I_1 and I_i are $p_1 = M_1 P$ and $p_i = M_i P$. A point p_1 detected in image I_1 is repeated in image I_i if the corresponding point p_i is detected in image I_i. To actually measure the repeatability, a relation between p_1 and p_i has to be established. In general this is impossible, but if the scene is planar this relation is defined by a homography : $p_i = H_{i1} p_1$.

The percentage of detected points which are repeated is the *repeatability rate*. A point is not in general detected exactly at position p_i, but rather in some neighbourhood of p_i. The size of this neighbourhood is denoted by ϵ and repeatability within this neighbourhood is called *ϵ-repeatability*. Moreover, some points can not be repeated due to parts which are observed only in one of the images. Points d_1 and d_i which could potentially be detected in both images are defined by $\{d_1\} = \{p_1 \mid H_{i1} p_1 \in I_i\}$ and $\{d_i\} = \{p_i \mid H_{1i} p_i \in I_1\}$ where $\{p_1\}$ and $\{p_i\}$ are the points detected in images I_1 and I_i. The set of point pairs (d_i, d_1) which correspond within an ϵ-neighbourhood is $D(\epsilon) = \{(d_i, d_1) \mid dist(d_i, H_{i1} d_1) < \epsilon\}$. The repeatability rate $r_i(\epsilon)$ for image I_i is then defined by :

$$r_i(\epsilon) = \frac{|D(\epsilon)|}{\min(n_i, n_1)}$$

where $n_1 = |\{d_1\}|$ and $n_i = |\{d_i\}|$. We can easily verify that $0 \leq r_i(\epsilon) \leq 1$.

3.2 Experimental Conditions

The repeatability rate of the different interest point detectors has been evaluated under various transformations : image rotation, scale change, variation of illumination and viewpoint change. Two different types of illumination are considered: a uniform variation and a complex variation. Stability to image noise has also been tested.

We have used two different planar scenes for our experimentation: "Van Gogh" and "Asterix". The "Van Gogh" scene is the "sower" painting shown in figure 1. The "Asterix" scene can be seen in figure 3. The two scenes are very different. The "Van Gogh" scene contains a lot of texture and the "Asterix" scene mostly line drawings.

To obtain an accurate repeatability rate, the computation of the homography has to be precise and independent of the detected points. Independent, subpixel localization is required. We arrange this by taking a second image for each position of the camera, one with black dots projected onto the scene (cf. figure 3). The black dots are extracted very precisely by fitting a template. The centres of these dots are then used to compute the homography. A least median square method makes the computation of the homography robust. While recording the sequence, the scene and the projection mechanism (overhead projector) remain fixed.

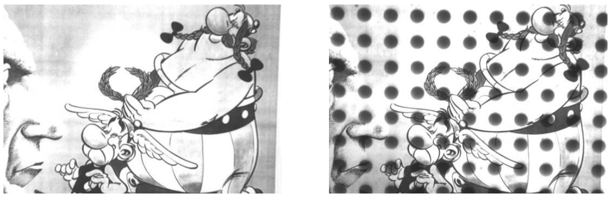

Figure 3: Two images of the "Asterix" scene. On the left the image of the scene, on the right the image with black dots projected onto the scene.

3.3 Results for repeatability

We first compare the two different versions of Harris. The one with better results is included in the comparison of different detectors. Due to space limitations we present only the results for the "Van Gogh" sequences; results for the "Asterix" sequences are given in [11]. The results for these two sequences are very similar; the "Asterix" scene confirms the results presented here.

3.3.1 Comparison of the two Harris detectors

Figure 4 compares the two different versions of the Harris detector in the presence of image rotation (graph on the left) and scale change (graph on the right). We see that the repeatability of the OurHarris version is better in both cases. The results of the standard version vary with image rotation, with the worst results being obtained for an angle of 45°. This is due to the fact that the standard version uses non-isotropic discrete filters. In conclusion, a stable implementation of the derivatives significantly improves the repeatability of the Harris detector.

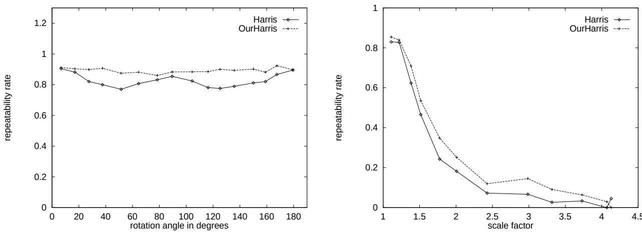

Figure 4: Comparison of Harris and OurHarris. On the left the repeatability rate for an image rotation and on the right the repeatability rate for a scale change. $\epsilon = 1.5$ pixels.

3.3.2 Image rotation

Image rotations are obtained by rotating the camera around its optical axis using a special fixation mechanism. Figures 1 and 2 show two images of the rotation sequence between which the image rotation is 155°. The repeatability rate for the rotation sequence is displayed in figure 5. The rotation angles vary between 0° and 180°. The graph on the left displays the repeatability rate for a localization error ϵ of 0.5 pixels which corresponds to a precision of one pixel. The graph on the right shows the results for an error of 1.5 pixels which signifies that the detected point lies in the neighbourhood of the predicted point.

For both localization errors the OurHarris detector gives the best results; results are invariant to rotation. At $\epsilon = 1.5$ the repeatability rate of OurHarris is almost 100%. Computing Harris only on the image contours (Cottier) makes the results worse. This is due to an additional instability in the contour extraction. The Heitger detector is not invariant to rotation, as it uses derivatives computed in several fixed directions. Its results are worst at a rotation of 90° which is due to the chosen filter directions. The worst results are obtained by the method based on the intersection of line segments (Horaud). This can be explained by the fact that contour extraction, line detection and intersection of lines introduce cumulative errors.

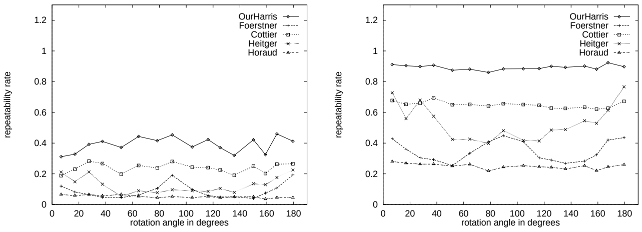

Figure 5: Repeatability rate for image rotation. $\epsilon = 0.5$ for the left graph and $\epsilon = 1.5$ for the right graph.

3.3.3 Scale change

Scale change is investigated by varying the focal length of the camera. Figures 1 and 2 show two images of the sequence between which the scale factor is 1.4. Figure 6 shows the repeatability for the various detectors in the presence of scale changes. Evidently the detectors are very sensitive to scale changes. For an ϵ of 0.5 (1.5) repeatability is very poor for a scale factor above 1.5 (2). The OurHarris and Cottier detectors give the best results. The results of the other detectors are hardly usable. Above a scale factor of about 2.5, the results are mainly due to artifacts. At larger scales many more points are found due to the texture of the scene, so accidental correspondences are more likely.

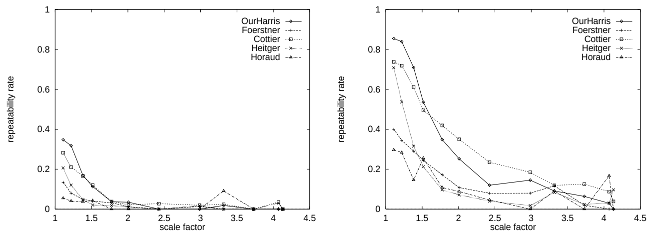

Figure 6: Repeatability rate for scale change. $\epsilon = 0.5$ for the left graph and $\epsilon = 1.5$ for the right graph.

3.3.4 Illumination variation

Uniform illumination variation

Uniform illumination variation is obtained by changing the camera aperture. To quantify the change we use *relative greylevel* - the ratio of mean greylevel of an image of the sequence to that of the reference image. Figure 7 shows images of the sequence.

Figure 7: Uniform illumination variation : from left to right images with relative greylevel 0.6, 1 and 1.7.

Figure 8 displays the results for a uniform illumination variation. Even for a relative greylevel of 1 there is not 100% repeatability due to image noise (two images of a relative greylevel of 1 have been taken, one reference image and one test image). In both graphs, the repeatability decreases smoothly in proportion to the relative greylevel. OurHarris and Heitger obtain better results than the other detectors ; the OurHarris detector gives slightly better results than Heitger.

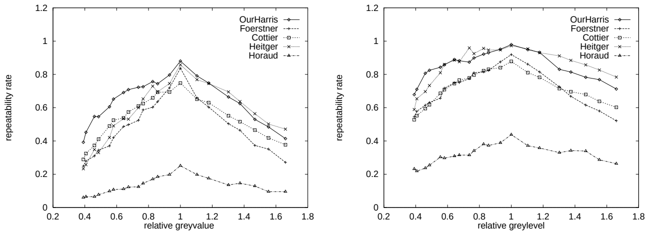

Figure 8: Repeatability rate for uniform illumination variations. $\epsilon = 0.5$ for the left graph and $\epsilon = 1.5$ for the right graph.

Complex illumination variation

We have also considered more complex illumination variations in which the light source moves in an arc from approximately $-45°$ to $45°$. Figure 9 shows images of the sequence.

Figure 9: Complex illumination variation : from left to right image 0 (reference image), 6 and 11. The light source is rightmost for image 0. For image 6 the light source is in front of the scene. The light source is leftmost for image 11.

Figure 10 displays the repeatability results. OurHarris is better than the other detectors. For an ϵ of 1.5 results are not modified by a complex illumination variation. Illumination direction has little effect as the computations of the detectors are local.

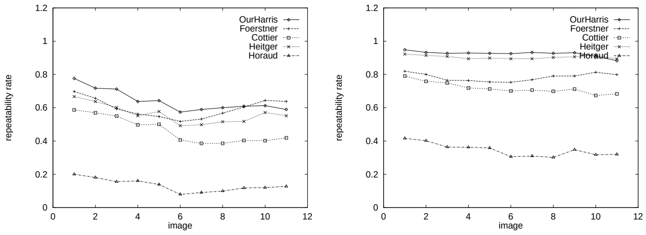

Figure 10: Repeatability rate for complex illumination variation. $\epsilon = 0.5$ for the left graph and $\epsilon = 1.5$ for the right graph.

3.3.5 Viewpoint Change

To measure repeatability in the presence of viewpoint changes, the position of the camera is moved in an arc around the scene. The angle varies from approximately $-50°$ to $50°$. The different viewpoints are approximately regularly spaced. Figure 11 shows images of the sequence.

Figure 12 displays the results for a viewpoint change. The OurHarris detector gives results superior to those of the other detectors. The results degrade

Figure 11: Viewpoint change sequence: from left to right image 0, 7 (reference image) and 15. Image 0 is taken at the rightmost position of the camera. For image 7 the camera is in front of the painting. Image 15 is taken at the leftmost position.

rapidly for $\epsilon = 0.5$, but significantly more slowly for $\epsilon = 1.5$. For this ϵ the repeatability of OurHarris is always above 60% except for image 0. These results show a good repeatability in the presence of perspective deformations.

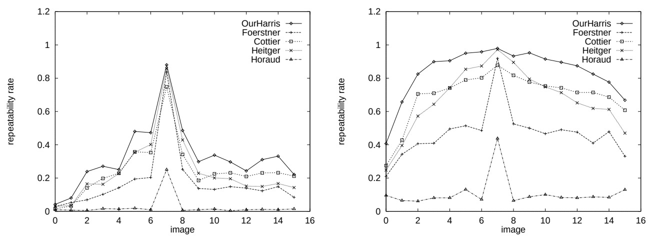

Figure 12: Repeatability rate for viewpoint changes. $\epsilon = 0.5$ for the left graph and $\epsilon = 1.5$ for the right graph.

3.3.6 Camera noise

To study repeatability in the presence of image noise, a static scene has been recorded several times. The results of this experiment are displayed in figure 13. We can see that all detectors give good results except the Horaud one. OurHarris gives the best results, followed closely by Heitger. For $\epsilon = 1.5$ these two detectors obtain a rate of nearly 100%.

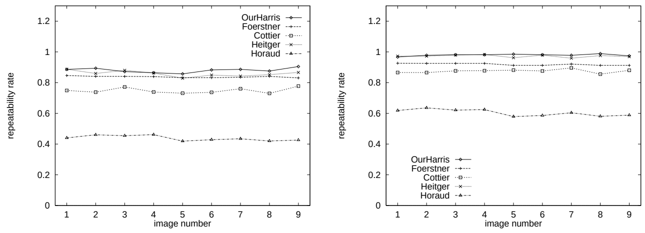

Figure 13: Repeatability rate for camera noise. $\epsilon = 0.5$ for the left graph and $\epsilon = 1.5$ for the right graph.

4 Information content

4.1 Entropy

This section gives a short introduction to entropy. Entropy measures the randomness of a variable. The more random a variable is the bigger the entropy. In the following we are not going to deal with continuous variables, but with partitions [9]. Any continuous variable can be partitioned by ranging its values into intervals. The entropy of a partition $\{\mathcal{A}\}$ is $H(\mathcal{A}) = -\sum_i p_i \log(p_i)$ where p_i is the probability of $\mathcal{A}_i$. If $\mathcal{B}$ is a new partition formed by subdivisions of the sets of $\mathcal{A}$ then $H(\mathcal{B}) > H(\mathcal{A})$. The size of the partition influences the result.

Entropy measures average information content. In information theory the information content of a message i is defined as $I = log(1/p_i) = -log(p_i)$. The information content of a message is inversely related to its probability. If $p_i = 1$ the event always occurs and no information is attributed to it: $I = 0$. The average information content per message of a set of messages is then $-\sum_i p_i \log(p_i)$ which is its entropy.

In the case of interest points we would like to know how much average information content an interest point "transmits", as measured by its greylevel pattern.The more distinctive the greylevel patterns are, the larger is the entropy.

4.2 Characterization of local patterns

To measure the distribution of local greyvalue patterns at interest points, we have to define a measure which describes such a pattern. Collecting unordered pixel values at an interest point is not useful since such a collection does not represent the shape of the signal around the point. Collecting ordered pixel values (eg. from left to right and from top to bottom) respects the shape but is not invariant to rotation. We have therefore chosen to characterize interest points by local greyvalue rotation invariants which are combinations of derivatives. Derivatives up to Nth order describe the intensity function locally up to that order. They are computed stably by convolution with Gaussian derivatives. This set of derivatives is called the "local jet" [8]. The "local jet" of order N at a point $\mathbf{x} = (x, y)$ for image I and scale σ is defined by $J^N[I](\mathbf{x},\sigma) = \{L_{i_1 \ldots i_n}(\mathbf{x},\sigma) \mid (\mathbf{x},\sigma) \in I \times \mathbb{R}^+ ; n = 0, \ldots, N\}$ in which $L_{i_1 \ldots i_n}(\mathbf{x},\sigma)$ is the convolution of image I with the Gaussian derivatives $G_{i_1 \ldots i_n}(\mathbf{x},\sigma)$ and $i_k \in \{x, y\}$.

In order to obtain invariance under the group $SO(2)$ (rigid rotations in the image), [8, 10] compute differential invariants from the "local jet". The invariants used in this work are

$$\vec{\mathcal{V}}[0..3] = \begin{bmatrix} L_x L_x + L_y L_y \\ L_{xx} L_x L_x + 2 L_{xy} L_x L_y + L_{yy} L_y L_y \\ L_{xx} + L_{yy} \\ L_{xx} L_{xx} + 2 L_{xy} L_{xy} + L_{yy} L_{yy} \end{bmatrix} \qquad (1)$$

The computation of entropy requires a partitioning of the space of $\vec{\mathcal{V}}$. Normalization allows us to use the same partition in all dimensions of $\vec{\mathcal{V}}$.

The normalization is based on Mahalanobis distance. Two random variables are compared by $d_M(\vec{\mathcal{V}}_2, \vec{\mathcal{V}}_1) = \sqrt{(\vec{\mathcal{V}}_2 - \vec{\mathcal{V}}_1)^T \Lambda^{-1} (\vec{\mathcal{V}}_2 - \vec{\mathcal{V}}_1)}$ where Λ is the covariance of $\vec{\mathcal{V}}$. The covariance matrix Λ corresponds to the uncertainty due to the noise variance of the data. This matrix Λ is symmetric definite positive and can be decomposed : $\Lambda^{-1} = P^T D P$ where D is diagonal and P an orthogonal matrix. Normalized vectors are then obtained by $\sqrt{D} P \vec{\mathcal{V}}$.

4.3 Results for information content

In this section, we compute the entropy for the detectors compared. Random points are included in our comparison : for each image we compute the mean number m of interest points extracted by the different detectors. We then select m random points over the image using a uniform distribution. To obtain a significant measure of entropy, we use a set of 1000 images. These images are of different types : aerial images, images of paintings and images of toy objects. For each detector we compute the interest points for the set of images and characterize each point by a vector of local greylevel invariants. Invariants are then normalized and the entropy of the distribution is computed.

The threshold used for partitioning the invariants is 20. The σ used for computing the greylevel invariants is 3. The results ordered according to entropy are presented in table 1. It shows that the detector OurHarris has the highest entropy. The results obtained for Heitger are almost as good. The two detectors based on line extraction obtain worse results. This can be explained by their limitation to contour lines which reduces their variety and thus the entropy. Unsurprisingly, the results obtained for all of the interest point detectors are significantly better than those for random points. The difference between the results of OurHarris and random points is about a factor of two.

detector	entropy
OurHarris	6.049526
Heitger	5.940877
Horaud	5.433776
Cottier	4.846409
Förstner	4.523368
Random	3.300863

Table 1: The information content for different detectors.

5 Conclusion

In this paper we have presented a comparison and evaluation of several interest point detectors based on two criteria : repeatability and information content. Repeatability was evaluated in the presence of different imaging conditions. In all cases the results of the OurHarris detector are better than or equivalent to those of the other detectors. For this detector, interest points are largely independent of the imaging conditions ; points are geometrically stable. Results for information content show that OurHarris obtains the best results. All of the detectors have significantly more information content than randomly selected points, so they do manage to select "interesting" points.

The framework defined in this paper allows the quantitative comparison of new algorithms for interest point extraction to existing ones. One possible extension is to define similar frameworks for other low-level features. Another extension would be to design an improved detector with respect to the two evaluation criteria. Concerning repeatability, we have seen that detectors show rapid degradation in the presence of scale change. To solve this problem, detectors could be included into a multi-scale framework. As this is quite time-consuming, another solution might be to estimate the scale at which the best results would be obtained. Concerning information content, we think that studying which kinds of greylevel patterns occur frequently and which ones are rare, will help us to design a detector with even higher information content.

References

[1] P. Brand and R. Mohr. Accuracy in image measure. *Videometrics III*, volume 2350, pages 218–228, 1994.

[2] J.C. Cottier. Extraction et appariements robustes des points d'intérêt de deux images non etalonnées. Internal Report, 1994.

[3] R. Deriche. Recursively implementing the Gaussian and its derivatives. Technical report, INRIA, 1993.

[4] W. Förstner. A framework for low level feature extraction. In *ECCV*, pages 383–394, 1994.

[5] C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey Vision Conference*, 1988.

[6] F. Heitger, L. Rosenthaler, R. von der Heydt, E. Peterhans, and O. Kuebler. Simulation of neural contour mechanism: from simple to end-stopped cells. In *Vision Research*, 32(5):963–981, 1992.

[7] R. Horaud, T. Skordas, and F. Veillon. Finding geometric and relational structures in an image. In *ECCV*, pages 374–384, 1990.

[8] J.J. Koenderink and A.J. van Doorn. Representation of local geometry in the visual system. In *Biological Cybernetics*, 55:367–375, 1987.

[9] A. Papoulis. *Probability, Random Variables, and Stochastic Processes.* McGraw-Hill, Inc., 1991.

[10] B.M Romeny, L.M.J Florack, A.H Salden, and M.A Viergever. Higher order differential structure of images. In *IVC*, 12(6):317–325, 1994.

[11] C. Schmid, R. Mohr, and C. Bauckhage. Comparing and evaluating interest points. Extended Version, 1997.
ftp.inrialpes.fr/pub/movi/publications/schmid_iccv98_ext.ps.